Urban Insecurity and Public Security Expenditures Data (in Progress)

Gulay Gunluk-Senesen
Istanbul University
Faculty of Political Sciences
Istanbul – Turkey
gulaygs@istanbul.edu.tr

Yasin Kutuk
Altınbaş University
School of Economics, Administrative and Social Sciences
Istanbul – Turkey
yasinkutuk@gmail.com

DATA DISPOSITION MANUAL

This study utilizes two datasets simultaneously. The first of these is "Global Database of Events, Language, and Tone" (GDELT hereafter) database, and the other is Integrated Crisis Early Warning System (abbreviated as ICEWS) database.

GDELT DATA

GDELT covers most of the events starting from 1979 to present, which is issued in a Cloud Platform 1 2 due to its hugeness, that is, it can be easily regarded as a big data. GDELT from the cloud platform is utilized by selecting its full event database sub database, specifically the steps sequentially are "gdelt-bq > full > events" which covers the first version of GDELT. The part this study mostly focuses is restricted with the events in which Turkey (encoded as firstly "TUR" in general 3 in "Actor1Code" or "Actor2Code") is stated as Source (Actor1) or Target (Actor2) in order to cover all events where Turkey is an active or inactive player.

This dataset was mined by following SQL command structure to select Turkey including it as active or inactive player in the events:

[1] https://console.cloud.google.com/bigquery?project=gdelt-bq
[2] This dataset were mined by SQL command structure to select Turkey including it as active or inactive player in the events in Appendix A.
[3] Actually, there are so many TUR encodings following the first. The best appropriate one is only "TUR", the other example encodings also specify the type of event, such as TURMIL (for military events), TURGOV (for governmental events), TURCVL, TURJUD etc.. This study considers only the first three letters to cover all events in Turkey.

Table 1. SQL Command

```
SELECT
  *
FROM
  `gdelt-bq.full.events`
WHERE
  Year=i AND (Actor1CountryCode ='TUR' OR Actor2CountryCode ='TUR')
ORDER BY
  GLOBALEVENTID
```

where i = {2006, ..., 2018} that specifies the year of event in Table 1. Since the dataset do not allow us to download it for a specific range due to its size, i was repeated for each respective year to make sure that its size was reduced to downloadable level.

ICEWS DATA

Another database named Integrated Crisis Early Warning System (abbreviated as ICEWS) led by O'brien (2010) to generate an alert system and to predict possible ones for future political events by monitoring political and social events. Boschee et al. (2015) issued ICEWS on Harvard's Dataverse 4 covering events for the years between 1995 and 2020 on a yearly basis (Shilliday and Lautenschlager, 2012).

It compiled the events more than 20 millions of news, articles in many languages and classified them according to CAMEO system. This paper uses the same procedure again what steps were done in GDELT. Yearly tab delimited datasets were downloaded from Harvard's Dataverse and were column-wise combined to match the variables throughout the years between 2006 and 2018. Turkey was selected again as active (Source) or passive (Target) player in the events. Later, the events where its latitude and longitude information is outside of Turkey except the cross border provinces/cities within neighboring countries of Turkey by following the Table 3 were removed. In addition to this, those provinces were accounted into the cross-border provinces of Turkey. After that, the events which are encoded using CAMEO event classification system were filtered in accordance with the scope of this study by

[4] https://dataverse.harvard.edu/dataverse/icews

utilizing Table 3 again. The datasets used and downloaded from ICEWS' repository can be tracked by following the Table 2.

Table 2. ICEWS Data sources

Name of Dataset	MD5 Code
events.2006.20150313083752.tab.zip	MD5: 8da71a3349bcea3391f9f8b1a77f4545
events.2007.20150313083959.tab.zip	MD5: d827c080883b523f9687482a7ea357af
events.2008.20150313084156.tab.zip	MD5: 35dde1eb434c054463acf51133eac6b7
events.2009.20150313084349.tab.zip	MD5: 8bb053ac42d91573a12f8b7805d658c1
events.2010.20150313084533.tab.zip	MD5: 1358b90c3f8e588c19844c0b190c679e
events.2011.20150313084656.tab.zip	MD5: 9121c54145b81332b89e336c323ee4f6
events.2012.20150313084811.tab.zip	MD5: 26a49b6967c79e5f4a78d344ec4d1520
events.2013.20150313084929.tab.zip	MD5: 2404025df02ec222feb689369fe190f2
events.2014.20160121105408.tab	MD5: bca3cc7f31d026e5e810ba2cfc0a7d58
events.2015.20180710092545.tab	No MD5, 20 Variables, 955349 Observations - UNF:6:tnA+AIB7BPaZUt9NIjiO7w==
events.2016.20180710092843.tab	No MD5,20 Variables, 787612 Observations - UNF:6:dddIA+DxNaLjOXfN/uA/AA==
Events.2017.20200602.tab.zip	MD5: 8f1146036e78c1bc0fea545c6a30afa9
events.2018.20200427084805.tab	MD5: c860c2c30df2f029bdf5fb0858e95dfd
events.2019.20200427085336.tab	MD5: a12b4c456cdfa30010e49b7ca10adc1e

ICEWS was compiled to collect again, but now, for one variable, "Intensity". Intensity was measures in parallel with GoldsteinScale as of GDELT 5. It weighs only categories and subcategories of CAMEO Events ranging from -10 to +10, where 0 is again midpoint equalizer of this scale. Intensity was processed again into two ways, first aggregated them in sums in the base of 81 cities and years (from 2006 to 2018), second way calculated averages of all events occurred in respective city and year in the base of 81 cities and 13 years again. If

[5] https://github.com/conjugateprior/events/blob/master/inst/CAMEO.SCALE.txt .

this calculation returns as empty, this city with its certain year will get 0 as what has been done in GDELT aggregations.

Filtering Procedure

As it is previously stated, CAMEO is the main framework for a classification of events in a structure that is consistent and prevents a confliction between events (i.e. double counting), which is offered by Gerner et al. (2002) for widely accepted classification and taxonomy system for event data as a framework [6]. GDELT uses CAMEO event classification system with 20 main categories. This study with the scope of the events restricted GDELT database by selecting the codes presented under the variable named "EventCode" as follows:

Table 3. Selected CAMEO Verbs

No	Main Category	Event Codes (Selected)
1	Make Public Statement	
2	Appeal	
3	Express Intent To Cooperate	
4	Consult	
5	Engage In Diplomatic Cooperation	
6	Engage In Material Cooperation	62
7	Provide Aid	72,74
8	Yield	
9	Investigate	91,92,93
10	Demand	
11	Disapprove	1122
12	Reject	
13	Threaten	130,1322,1323,1324,133,134,138,1381, 1382,1383,1384
14	Protest	141,1411,1412,1413,1414,143,1431,1432,1433, 1434,1441,1442,1443,1444,145,1451,1452,1453,1454
15	Exhibit Force Posture	150,151,152,153,154
16	Reduce Relations	
17	Coerce	1712,172,1721,1722,1723,1724,173,175
18	Assault	180,181,182,1822,1823,183,1831,1832, 1833,1834,185,186
19	Fight	190,191,192,193,194,195,1951,1952
20	Use Unconventional Mass Violence	20,200,201,202,203,204,2041, 2042

[6] http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf

According to the Table 3, the numbers nearby its main category indicate Event Verb represented with the numbers which also show the levels (A detailed one is given in Table 4). Therefore, in addition to the main category (first level), Event Verbs are encoded with the second and third level to make it more specific. However, if the third level is specified under its second level, it makes a clear distinction between its second level not to allow double counting (in general, the second level for such case is stated as "Event Verb - not specified below" to prevent a double counting issue with its subcategory, third level). As it can be seen in the Table 3, the topics mostly covered Fight, Coerce and Use Unconventional Mass Violence, which are the most probable candidates to break public safety and to increase defence expenditures.

In addition to above restriction, GDELT data is also filtered by where the event occured. To expand, it is known that the data is related to Turkey for some specific events. However, the main purpose is actually to distribute these events to cities of Turkey. To do so, "ActionGeo_FullName" was employed to catch where the event was occured, which the variable incudes the information of the place of event in village, municipality, province/city, territory or again the country level. Frankly, this variable does not show consistency in terms of the place of event. To distribute correctly the events to cities of Turkey, two ways were followed:

- Villages, municipalities are mined to count them according to where they belong to within the border of a city in Turkey in detail. This study does not stop there, it also takes into account two variables "ActionGeo_Lat" consisting of latitude information of the event and "ActionGeo_Long" having longitude information of the event. If any of them exceeds the latitude and longitude informations of Turkey, the event is a candidate of a deletion.
- The event might be occurred outside of Turkey, however, it might be related to provinces which are on the border of Turkey. Before deletion of the candidate event, such events were counted as if they were neighbouring city of Turkey. A detailed counting process for the cities of Turkey is provided in Table 5 in Appendix.
- Some news were repeatedly given in GDELT. It is examined in detail for two variables to prevent double counting to capture them: "GLOBALEVENTID" that is a special code for each row and contains 9 digits unique number, and "SOURCEURL" that provides the

source hyperlink which was crawled to construct event text. GLOBALEVENTID shows some clue with its first 8 digits for a repetition in events, so, if there was a duplication in there, their respective rows were removed. If there was another repetition for SOURCEURL, their rows were again deleted to prevent double counting for a certain event.

After all these filtering and mining processes, GDELT database was utilized by considering its four basic variables, which are collapsed into years and 81 provinces of Turkey for the time-span between 2006 and 2018. The first one is "AvgTone" which calculates a semantic score for a certain event ranging from -10 to +10, where 0 plays a role as midpoint equalizer. In "AvgTone", the content of events was also scored and were weighted this variable, i.e., two protests which have 150 people and 1.5 billion people have different AvgTone scores. The second one is "GoldsteinScale" [7] that classifies CAMEO codes directly with their respective categories in the range of -10 and +10 again. Nonetheless, it is lack of weighting of the content of event since it is a measure free of how many people were involved in the event. The explanations of variables can be found on the Cloud Platform or other sources [8]. AvgTone and GoldsteinScale variables were aggregated in two ways: summing and averaging. The sums of AvgTone and GoldsteinScale in the bases of years and 81 cities were calculated to indicate the Intensity Sum of the events occurred in respective city for certain year to measure perception of safety. Again, The average of AvgTone and GoldsteinScale in the same bases were also calculated to get a general picture of perception of safety respective city for its year. If there was no event for a city for respective year, it yielded as 0.

[7] https://www.gdeltproject.org/data/lookups/CAMEO.goldsteinscale.txt

[8] https://github.com/linwoodc3/gdelt2HeaderRows/blob/master/schema_csvs/GDELT_2.0_Events_Column_Labels_Header_Row_Sep2016.csv

Table 4. Full List of Selected CAMEO Events

Code	Event Verb	Code	Event Verb
6	*ENGAGE IN MATERIAL COOPERATION*	15	*EXHIBIT MILITARY POSTURE*
62	Cooperate militarily	150	Demonstrate military or police power- not specified below
7	*PROVIDE AID*	151	Increase police alert status
72	Provide military aid	152	Increase military alert status
74	Provide military protection or peacekeeping	153	Mobilize or increase police power
9	*INVESTIGATE*	154	Mobilize or increase armed forces
91	Investigate crime, corruption	17	*COERCE*
92	Investigate human rights abuses	1712	Destroy property
93	Investigate military action	172	Impose administrative sanctions- not specified below
11	*DISAPPROVE*	1721	Impose restrictions on political freedoms
1122	Accuse of human rights abuses	1722	Ban political parties or politicians
13	*THREATEN*	1723	Impose curfew
130	Threaten- not specified below	1724	Impose state of emergency or martial law
1322	Threaten to ban political parties or politicians	173	Arrest, detain, or charge with legal action
1323	Threaten to impose curfew	175	Use tactics of violent repression
1324	Threaten to impose state of emergency or martial law	18	*ASSAULT*
133	Threaten with political dissent, protest	180	Use unconventional violence- not specified below
134	Threaten to halt negotiations	181	Abduct, hijack, or take hostage
138	Threaten with military force- not specified below	182	Physically assault- not specified below
1381	Threaten blockade	1822	Torture
1382	Threaten occupation	1823	Kill by physical assault
1383	Threaten unconventional violence	183	Conduct suicide, car, or other non-military bombing- not specified below
1384	Threaten conventional attack	1831	Carry out suicide bombing
14	*PROTEST*	1832	Carry out vehicular bombing
141	Demonstrate or rally- not specified below	1833	Carry out roadside bombing
1411	Demonstrate for leadership change	1834	Carry out location bombing
1412	Demonstrate for policy change	185	Attempt to assassinate
1413	Demonstrate for rights	186	Assassinate
1414	Demonstrate for change in institutions, regime	19	*FIGHT*
143	Conduct strike or boycott- not specified below	190	Use conventional military force- not specified below
1431	Conduct strike or boycott for leadership change	191	Impose blockade, restrict movement
1432	Conduct strike or boycott for policy change	192	Occupy territory
1433	Conduct strike or boycott for rights	193	Fight with small arms and light weapons
1434	Conduct strike or boycott for change in institutions, regime	194	Fight with artillery and tanks
1441	Obstruct passage to demand leadership change	195	Employ aerial weapons- not specified below
1442	Obstruct passage to demand policy change	1951	Employ precision-guided aerial munitions
1443	Obstruct passage to demand rights	1952	Employ remotely piloted aerial munitions
1444	Obstruct passage to demand change in institutions, regime	20	*USE UNCONVENTIONAL MASS VIOLENCE*
145	Protest violently, riot- not specified below	200	Use unconventional mass violence- not specified below
1451	Engage in violent protest for leadership change	201	Engage in mass expulsion
1452	Engage in violent protest for policy change	202	Engage in mass killings
1453	Engage in violent protest for rights	203	Engage in ethnic cleansing
1454	Engage in violent protest for change in institutions, regime	204	Use weapons of mass destruction- not specified below
		2041	Use chemical, biological, or radiological weapons
		2042	Detonate nuclear weapons

Table 5. Neighbouring provinces of Turkey border for events counted in Turkey

Country	**Provinces in Border**			**Province in Turkey**
Greece	Kastanies (Horozlu)	Pythio (Kuleliburgaz)	Kipoi (Turkish	Edirne
Bulgaria	Khaskovo	Yambol		Edirne
	Burgas			Kırklareli
Syria	Latakia	Idlib		Hatay
	Aleppo			Kilis, Gaziantep
	Ar Raqqah, Raqqah			Gaziantep
	Hasakah			Sanliurfa, Mardin, Sirnak
Iraq	Muhafazat Ninawa	Dihok		Şırnak
	Muhafazat Arbil	Kurdistan		Hakkari
Iran	Urmiye			Hakkari, Van
Armenia	Ararat	Armavir		Igdir
	Aragatsavan	Shirak		Kars
Georgia	Samtshe	Cavaheti	Adjara	Artvin